

Appl. No. 10/784,334  
Amdt. dated March 23, 2005

**Amendments to the Claims:**

This listing of claims will replace all prior versions, and listings, of claims in the application:

1. – 21. (Cancelled)

22. (New) A method for sizing content on an information processing system for language translation, the method comprising:

retrieving content containing markup tags, the content containing markup tags being viewable by a browser and being in a first language;

parsing the content containing markup tags into a plurality of translatable components;

generating a respective unique identifier for at least one respective translatable component within the plurality of translatable components of the content containing markup tags, the respective unique identifier identifying one respective translatable component within the plurality of translatable components;

storing the plurality of translatable components and corresponding unique identifiers for each translatable component within the plurality of translatable components; and

generating statistics based on the plurality of translatable components.

23. (New) The method of claim 22, further comprising:

queuing the plurality of translatable components and corresponding unique identifiers for translation of the plurality of translatable components into a second language.

24. (New) The method of claim 22, wherein the retrieving content containing markup tags comprises:

retrieving all content containing markup tags from a data source in a first language.

Appl. No. 10/784,334  
Amdt. dated March 23, 2005

25. (New) The method of claim 22, wherein translatable components within the plurality of translatable components include any one of:
- a text segment;
  - an image file linked from the content containing markup tags with text to be translated;
  - a multimedia file, linked from the content containing markup tags, with text or audio to be translated;
  - a file, linked from the content containing markup tags, with text to be translated;
  - a file, linked from the content containing markup tags, with image with text to be translated;
  - a file, linked from the content containing markup tags, with audio to be translated;
- and
- a file, linked from the content containing markup tags, with video with at least one of text and audio to be translated.
26. (New) The method of claim 22, wherein the generating a respective unique identifier further comprises:
- generating a unique identifier for each of the plurality of translatable components that is a text segment, wherein the unique identifier is any one of a hash code, a checksum and a mathematical algorithm for generating a unique identifier based on one or more of the segments.
27. (New) The method of claim 22, wherein the storing the plurality of translatable components comprises:
- storing the plurality of translatable components and corresponding unique identifiers such that translatable components within the plurality of translatable components are accessed using a corresponding unique identifier.
28. (New) The method of claim 22, wherein the statistics comprise at least one of:
- a file count;

Appl. No. 10/784,334  
Amdt. dated March 23, 2005

- a page count;
- a text segment count;
- a unique text segment count;
- a unique text segment word count; and
- a word count.

29. (New) The method of claim 23, wherein the queuing the plurality of translatable components comprises:

- queuing the plurality of translatable components and corresponding unique identifiers for human translation of the plurality of translatable components into a second language.

30. (New) The method of claim 29, wherein the first language includes any one of English, French, Spanish, German, Portuguese, Italian, Japanese, Chinese, Korean, and Arabic and the second language is different than the first language and includes any one of English, French, Spanish, German, Portuguese, Italian, Japanese, Chinese, Korean, and Arabic.

31. (New) The method of claim 22, further comprising:  
generating a web page having a link to each file of the web site.

32. (New) The method of claim 22, wherein the parsing is based upon the markup tags.

33. (New) A computer system for sizing content on an information processing system for language translation, the computer system comprising:

- a connection for retrieving content containing markup tags, the content containing markup tags being viewable by a browser and being in a first language;

- a processor for:

- parsing the content containing markup tags into a plurality of translatable components; and

Appl. No. 10/784,334  
Amdt. dated March 23, 2005

generating a respective unique identifier for at least one respective translatable component within the plurality of translatable components of the content containing markup tags, the respective unique identifier identifying one respective translatable component within the plurality of translatable components; and

a database for storing the plurality of translatable components and corresponding unique identifiers for each translatable component within the plurality of translatable components; and

a database management system for generating statistics based on the plurality of translatable components.

34. (New) The computer system of claim 33, wherein the database management system further queues the plurality of translatable components and corresponding unique identifiers for translation of the plurality of translatable components into a second language.

35. (New) The computer system of claim 33, wherein the connection is further for retrieving all content from a data source in a first language

36. (New) The computer system of claim 33, wherein translatable components within the plurality of translatable components include any one of:

a text segment;

an image file, linked from the content containing markup tags, with text to be translated;

a multimedia file, linked from the content containing markup tags, with text or audio to be translated;

a file, linked from the content containing markup tags, with text to be translated;

a file, linked from the content containing markup tags, with image with text to be translated;

a file, linked from the content containing markup tags, with audio to be translated;

and

Appl. No. 10/784,334  
Amdt. dated March 23, 2005

a file, linked from the content containing markup tags, with video with at least one of text and audio to be translated.

37. (New) The computer system of claim 33, wherein the generating a respective unique identifier further comprises:

generating a unique identifier for each of the plurality of translatable components that is a text segment, wherein the unique identifier is any one of a hash code, a checksum and a mathematical algorithm for generating a unique identifier based on one or more of the segments.

38. (New) The computer system of claim 33, wherein the database is further for storing the plurality of translatable components and corresponding unique identifiers such that a translatable component may be accessed using a corresponding unique identifier.

39. (New) The computer system of claim 33, wherein the statistics comprise at least one of:

- a file count;
- a page count;
- a unique segment count;
- a unique text segment count;
- a unique text segment word count; and
- a word count.

40. (New) The computer system of claim 33, wherein the database management system further queues the plurality of translatable components and corresponding unique identifiers for human translation of the plurality of translatable components into a second language.

41. (New) The computer system of claim 40, wherein the first language includes any one of English, French, Spanish, German, Portuguese, Italian, Japanese, Chinese, Korean, and Arabic and the second language is different than the first language and includes any

Appl. No. 10/784,334  
Amdt. dated March 23, 2005

one of English, French, Spanish, German, Portuguese, Italian, Japanese, Chinese, Korean, and Arabic.

42. (New) The computer system of claim 33, wherein the database management system further generates a web page having a link to each file of the web site.

43. (New) The computer system of claim 33, wherein the processor parses the content based upon the markup tags.

44. (New) A computer program product including computer instructions for sizing content on an information processing system for language translation, the computer instructions including instructions for:

- retrieving content containing markup tags, the content containing markup tags being viewable by a browser and being in a first language;

- parsing the content containing markup tags into a plurality of translatable components;

- generating a respective unique identifier for at least one respective translatable component within the plurality of translatable components of the content containing markup tags, the respective unique identifier identifying one respective translatable component within the plurality of translatable components;

- storing the plurality of translatable components and corresponding unique identifiers for each translatable component within the plurality of translatable components; and

- generating statistics based on the plurality of translatable components.

45. (New) The computer program product of claim 44, wherein the statistics comprise at least one of:

- a file count;

- a page count;

- a unique segment count;

- a unique text segment count;

Appl. No. 10/784,334  
Amdt. dated March 23, 2005

a unique text segment word count; and  
a word count.

46. (New) A method for sizing a web site for language translation, the method comprising:

retrieving content containing markup tags in a first language, the content containing markup tags being viewable by a browser;

parsing the content containing markup tags into a plurality of translatable components based upon markup tags contained in the content, wherein the translatable components are to be transmitted by human translators; and

generating statistics based on the translatable components retrieved from the web site.

47. (New) The method of claim 46, further comprising:

saving a cookie received from the web site; and

returning the cookie to the web site when requesting additional pages.

48. (New) The method of claim 46, wherein the content comprises an HTML form, and wherein the method further comprises:

filling out the HTML form with pre-defined information; and

submitting the HTML form with the predefined information to the web site.